CCS-3 Can Find 'Needles' in Data 'Haystacks'

Supercomputers make it possible today to generate huge data sets from hydrocodes. It is even possible to create visualizations from the data that can summarize what has occurred.

However, these visualizations are so simplified that they can fail to reveal problems buried deep within the masses of data. By analogy, consider trying to determine from a photograph whether one dime lies hidden in a gymnasium full of pennies.

Now, a team in the Computer and Computational Sciences Division (CCS) has developed a mathematical way to reach into these terabyte data sets and find relevant problems. The team includes John Hogden, Tony Warnock, Patricia Fasel, Richard Fortson, and Don Hush, all in the Modeling, Algorithms, and Informatics Group (CCS-3) of the Computer and Computational Sciences Division. The team also included Michael Cannon of CCS-3 until his recent retirement.

Their method is already being used in several interesting ways. They are, for example, assisting the Applied Physics Division (X Division) in checking for asymmetries in the computer simulation of imploding pits.

Their method can also be extremely useful in checking terabyte sets of data for transient events. They do this by checking an event to see how many standard deviations out from the mean it lies.

Cannon commented, "We find needles in haystacks!"

The team can also find bad values in mixed cells. Warnock has developed what Cannon called "a mathematical bump detector" that can find "bugs you never knew existed." These bad values would never be revealed in a visual tour of the data.

Consider one more application off the team's work: the ability to compare shapes and determine which ones are alike and which ones are different. The human eye and brain are very good at such discrimination, but in the past, it has been difficult to teach a computer to do it. Now, however, team members can use their mathematical methods to compare photos taken by PHERMEX (the pulsed high-energy radiographic machine emitting x-rays), for example. Their approach involves comparison of histograms of many lines from random point to random point on the surface.

Cannon noted that the method works in both two and three dimensions.

X Division, Physics Division (P Division), and the Dynamic Experimentation Division (DX Division) have expressed interest in this method of determining similarity and difference in shapes.

After a briefing on the team’s work, CCS Deputy Division Leader Stephen Lee commented, “This is very high-impact work with important ramifications in the nuclear weapons program and threat-reduction programs.”

(Taken from LALP-05-015, February issue)